APPLICATIONS OF DATA WAREHOUSING AS KNOWLEDGE POOL

Abobakr Aljuwaiber

Department of Administrative and Financial Sciences and Technologies, Applied College, Taibah University, Saudi Arabia

ABSTRACT

Increasing amounts of information and diverse formats have forced organizations to create large data repositories in response to the information explosion in the 21st century. As a result, a data warehouse is introduced as a large data repository. The purpose of this article is to describe the principles of data warehousing in business to enhance the generation of new knowledge throughout the organization. Some definitions also include methods of using data warehousing, namely query and data simulation. Then the steps before transforming the raw data into a data warehouse are considered. Project goals for data warehouses and metadata are discussed. This document aims to provide a clear and simple description of data warehousing terms and concepts, especially for busy managers and laymen. Therefore, they may only need basic and direct information about the data warehouses to gain complete understanding of the principles of the data warehouse.

KEYWORDS

Data warehouse, Business, Metadata, Knowledge pool.

1. INTRODUCTION

From the late mid-1990s, the data warehouses have become one of the crucial developments in the field of information systems. Data warehousing is an increasingly popular and powerful concept for applying information technology to solve business problems. In fact, a data warehouse improves the productivity of business decisions by consolidating, transforming, and integrating operational data, providing a consistent view of an organization. It can be used as a knowledge pool for the organizations where information and data can be retrieved and accessed [10]. Therefore, data warehouse is one of the knowledge sources. The data warehouse provides Online Analytical Processing (OLAP) tools. OLAP is described by Khan [8, p. 108] as “a business intelligence tool that addresses the need to perform multidimensional analysis”. It is one of the functions that aid knowledge workers (executives, managers, and analysts) in analysing data of business transactions stored in the dimensional data warehouse to make tactical and strategic business decisions [14, 15]. Hence, the data warehouse is the core of the analytics process and is the instrument that assures that the information extracted, transformed and loaded meets the quality standards that are necessary to have a quality output in the generation of organizational knowledge. This concept can assist knowledge workers with an intelligent analysis platform as a decision support system [12].

As new information and communication technologies emerged, companies needed to deal with big data analytics. Hence, data warehouses and database technologies were utilized to address this issue. Essentially, data warehousing assists to construct a huge repository of integrated data that is enhanced for analysis purposes [7]. A data warehouse has one of the most important
components, the metadata stored in the warehouse. Accordingly, the big challenge is how to extract meaningful value from the data set [5].

The specific purpose of this article is to explain the principles of the data warehousing in business which enhance the generation of new knowledge throughout the organization. This article also provides readers with the most important aspects of data storage and data usage. The paper considers various methods of transforming and integrating data warehouses and their real-world applications. The report begins with some definitions and uses of data warehouses. The next step is to convert raw data to a data warehouse. Then discuss the goals of the project. Finally, one of the most important components of a data warehouse, the so-called metadata, is described.

2. Definition of Data Warehouse

Data warehousing is a concept and not a product that can be brought off the shelf. Data warehouses are expensive to build and difficult to protect [16]. There are various definitions of a data warehouse; for instance, Rainardi [14, p. 1] defined a data warehouse as “a system that retrieves and consolidates data periodically from the source system into dimensional or normalized data store”. Thus, it can be viewed as a database that collects data from a variety of sources within an organization, including the company's processes, products, and customers [13]. Data warehouse collects briefed data from different sources, structured, and optimised for query access using OLAP (Online Analytical Processing) query tool. Chaudhuri and Dayal [3, p. 65] defined data warehousing as “a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, analyst) to make better and faster decisions”.

2.1. Why Data Warehousing?

The main idea of data warehousing comes from companies that have developed and implemented multiple information systems. Data warehouses are designed to store, extract, and transfer data for strategic analysis [6]. Furthermore, these processes help the data warehouse provide useful information to decision makers that aid to improve their business processes [11]. However, there is a bit of concern about security issues for a data warehouse that should be considered in the field of information systems since data warehouse is an open area (Gosain and Arora, 2015).

2.2. Using Data Warehouses

There are two basic ways to access information in a data warehouse which are queries and data mining.

1. Queries: Query can be defined as “the process of getting data from a data store, which satisfies certain criteria” [14, p. 11]. Data warehouse can be examined with query language such as Structured Query Language (SQL) or by using Online Analytical Processing (OLAP).

2. Data mining: Some research call data mining as knowledge discovery in database (KDD). Data mining is an experimental, exploratory, and iterative process that consists of a number of stages [17]. It helps end-users extract useful business information from a large database. Thus, data mining uses developed statistical processing or artificial intelligence programs to analyse the data and find patterns [6].
3. **Steps for Transforming Raw Data into a Data Warehouse**

Data warehouse converts raw data into a useful analytical tool for business decision-making. Some companies execute raw data into Online Transaction Processing (OLTP) systems, which follow day-to-day operations, for example, sales, purchases, and inventory change [14]. Furthermore, the integrated processes of transforming raw data into a data warehouse where various formats can be rigorously unified by the extract-transform-load (ETL) procedure into the database via extraction, consolidation, filtering, transformation, cleansing, conversion, and aggregation [4, 8, 9]. Figure 1 illustrates operations of getting data from original sources into the data warehousing processes, which are:

![Figure 1. Processes of transforming raw data into a data warehouse](image)

1. **Extraction:** This step usually deals with various data sources. In this step, the principle may take data out of its original source of the database. Then, transform it into a data warehouse infrastructure.
2. **Consolidation:** This means the process of integrating data from multiple sources into one database.
3. **Filtering:** In this process, we take the useful data and discard the rest.
4. **Transformation:** Involves subjecting the data to a number of operations before importing it and deriving new values to be loaded to the target system, or for validating the data from the source system.
5. **Cleansing:** This step is important to get a good quality of data that improves the accuracy of the data in the warehouse. However, a large volume of data in the warehouse may have errors or anomalies due to multiple sources involved [3], such as inconsistent field lengths, inconsistent descriptions, inconsistent value assignments, missing entries, and violation of integrity constraints. Hence, it is important that data in the warehouse needs to be correct.
6. **Conversion:** It means planning the raw data onto a new data field inside the warehouse data. Typing and translating the data into the format used by a warehouse.
7. **Aggregation:** In this last step many times the value of a data warehouse is in the abbreviated and extracted data it contains; the aggregation step classifies and joins data into useful metrics for analysis.

4. **Objectives Project of Data Warehouse**

The most important way to achieve the objectives of a project is to have team members that come from different backgrounds and have been trained in different disciplines. The team includes members with different academic backgrounds or major in different fields for example, some of
them being business specialist, technical systems specialists, data modellers, data analysts, and database administrators who are working to achieve the same goal [13].

In the meantime, team members try to get the response and attention of customer needs which is very important in global markets. When they know a customer’s desire this will help them to deal with every customer with individual care the customer demands [7, 13].

There are four objectives acceptable for almost any data-warehousing projects that are explained as follows:

- Develop a coordinated and collaborative information supply approach.
- Institute a consistent development approach and methodology for data warehousing.
- Ability the existing data warehouse experience and knowledge.
- Develop common terminology and concepts.

In addition to that, the better question may be is to find out if the team understands what information architecture is, and then ask whether or not the data warehouse project has one [13].

5. Metadata

Metadata can be defined as one of the most important phases of data warehousing. It is “data about data” that describes and is stored in the warehouse and has several uses [1]. Metadata is vital for building, maintaining, managing, and using a data warehouse. Moreover, it provides useful information for locating data stored in the data warehouse [2]. Metadata delivers some useful points including the following:

1. The location and description of the warehouse system and data components.
2. Names, definition, structure, and content of data warehouse and end observation.
3. Verification of imperious data sources.
4. For data to remain in the warehouse, consolidation and transformation rules are used, and these include mapping methods from the operational database.
5. Integration and transformation rules are used to deliver data to end-user analytical tools.
6. Data warehouse operational information which includes a history of warehouse updates, versions, and ownership authorizations [2, 8].

5.1. Metadata Trends

The trend towards integrating external data within data warehouses is clearly one of the best ways to model data warehouses. Thus, metadata trends have processes of integration of external and internal data into a warehouse that are illustrated in the following: (1) different data formats, (2) missing or invalid data, (3) levels have different aggregation, (4) indicative inconsistency, and (5) unknown data quality and timeliness. All these processes put additional load on the collection and management of the general metadata definitions [1, 8].

6. Conclusions

Increasing amounts of information and diverse formats have forced organizations to create large data repositories in response to the information explosion in the 21st century. As a result, a data warehouse is introduced as a large data repository. This article discussed in the beginning, definitions of data warehousing as the collection of data from multiple sources and making that data available to end users in a consolidated, consistent manner. It is essential that companies
improve multiple information systems in the data warehousing concept. Processes were also illustrated to convert raw data into a useful data warehouse, which includes extraction, consolidation, filtering, transformation, cleansing, conversion, and aggregation. A successful data warehouse project is the one that depends on the activities of the team members to achieve the goals and plans [13]. Metadata is one of the most important points of a data warehouse. The metadata trends are necessary to reduce the cost and increase the competitiveness of the business rapidly. Finally, it is important for businesses to consider the quick changes in information technology and communication and how to handle big data technology to solve the problems and changes in the modern data warehouse.

REFERENCES

Abobakr Aljuwaiber is an Associate Professor of Knowledge Management. He holds MBA and Ph. D. in knowledge management from University of East Anglia, UK. He is currently a Vice-Dean of Research and Consulting Institute at Taibah University and used to be a Head of Administrative and Financial Sciences and Technologies Department, and a Vice-Dean at Community College (Now as Applied College), Taibah University, Saudi Arabia. He has published several peer-reviewed researched papers in international journals and conference proceedings. His research interests focus on knowledge management, knowledge sharing, organisational learning, and communities of practice, strategic management.